KI-System Dokumentation für EU AI Act Compliance

1. Systemübersicht

1.1 Systembezeichnung

Name: licili - Kundenfeedback-Analyse-System

1.2 Zweck und Anwendungsbereich

- **Hauptfunktion:** Automatisierte Analyse und Zusammenfassung von Kundenfeedback
- **Zielgruppe:** Mitarbeiter des Unternehmens
- **Einsatzbereich:** Kundenfeedback-Management und -analyse

1.3 Systemarchitektur

Kundenfeedback Input → Preprocessing → KI-Analyse → Menschliche Überprüfung → Output

2. KI-Komponenten

2.1 Large Language Model (LLM)

- **Modell:** Meta Llama 3.3 70B
- **Hosting:** IONOS AI HUB (Deutschland/EU)
- **Funktion:** Generierung von Textzusammenfassungen und Lösungsvorschläge
- **Input:** Kundenfeedback-Texte
- **Output:** Strukturierte Zusammenfassungen und Lösungsvorschläge

2.2 Klassische ML-Modelle

- **Sentiment-Analyse:** Erkennung emotionaler Bewertungen
- **Themen-Extraktion:** Identifikation von Hauptthemen
- **Kontext-Analyse:** Einordnung des Feedback-Kontexts
- **Relevanz-Bewertung:** Priorisierung nach Wichtigkeit

2.3 Statistische Analyseverfahren

- Häufigkeitsanalysen
- Trend-Erkennung

- Kategorisierung

3. Datenverarbeitung

3.1 Input-Daten

- **Datentyp:** Kundenfeedback-Texte
- **Datenquellen:** Öffentliche und interne Quellen. Abhängig vom Hauptvertrag
- **Sensible Daten:** Keine biometrischen oder Gesundheitsdaten
- **Datenschutz:** Anonymisierung/Pseudonymisierung wo erforderlich

3.2 Verarbeitungsprozess

1. **Datenaufbereitung:** Bereinigung und Strukturierung
2. **KI-Analyse:** Automatische Verarbeitung durch LLM und ML-Modelle
3. **Qualitätskontrolle:** Automatische Plausibilitätsprüfung
4. **Output-Generierung:** Erstellung strukturierter Berichte

3.3 Output-Daten

- **Format:** Strukturierte Zusammenfassungen und Analysen
- **Kennzeichnung:** Klare Markierung als KI-generierte Inhalte
- **Bearbeitbarkeit:** Vollständige Editierbarkeit durch Mitarbeiter

4. Risikobewertung

4.1 AI Act Kategorisierung

- **Risikostufe:** Geringes Risiko
- **Begründung:**
    - Keine Hochrisiko-Anwendungsbereiche betroffen
    - Keine automatisierten Entscheidungen ohne menschliche Aufsicht
    - Nur interne Nutzung durch geschulte Mitarbeiter

4.2 Identifizierte Risiken

- **Bias in Sentiment-Analyse:** Mögliche Verzerrungen bei bestimmten Themen
- **Halluzinationen:** LLM könnte nicht vorhandene Inhalte generieren
- **Datenschutz:** Versehentliche Verarbeitung sensibler Daten

4.3 Risikominderungsmaßnahmen

- Regelmäßige Überprüfung der Output-Qualität
- Menschliche Validierung aller KI-Outputs
- Anonymisierung sensibler Daten vor Verarbeitung

5. Menschliche Aufsicht (Human-in-the-Loop)

5.1 Aufsichtsebenen

1. **Operative Ebene:** Mitarbeiter überprüfen und bearbeiten KI-Outputs
2. **Qualitätssicherung:** Stichprobenhafte Kontrolle der Analyseergebnisse
3. **Systemebene:** Monitoring der KI-Performance

5.2 Eingriffsmöglichkeiten

- **Vollständige Editierbarkeit:** Mitarbeiter können alle KI-Outputs ändern
- **Verwerfung:** Möglichkeit, KI-Outputs komplett zu verwerfen
- **Feedback-Loop:** Korrekturen fließen in Systemverbesserung ein

5.3 Schulung und Kompetenz

- Mitarbeiter sind über KI-Funktionsweise informiert
- Klare Prozesse für Qualitätskontrolle

6. Transparenz und Kennzeichnung

6.1 Nutzerinformation

- **Kennzeichnung:** Alle KI-generierten Inhalte sind als solche markiert
- **Erklärbarkeit:** Nutzer verstehen, wie Zusammenfassungen entstehen
- **Limitations:** Hinweise auf mögliche KI-Limitationen

6.2 Dokumentation für Endnutzer

- Benutzerhandbuch mit KI-Funktionsbeschreibung
- Hinweise zur Interpretation der Ergebnisse
- Kontaktmöglichkeiten bei Problemen

7. Technische Spezifikationen

7.1 Hosting und Infrastruktur

- **Anbieter:** IONOS AI HUB
- **Standort:** Deutschland/EU
- **Compliance:** DSGVO-konform

8. Governance und Compliance

8.1 Verantwortlichkeiten

- **System Owner:** Lucas Bäuerle, CTO
- **KI-Verantwortlicher:** Lucas Bäuerle, CTO

8.2 Prozesse

- **Änderungsmanagement:** Dokumentierte Prozesse für Systemupdates
- **Incident Management:** Vorgehen bei KI-Fehlern oder Problemen
- **Audit-Trail:** Nachverfolgbarkeit aller Systemänderungen

8.3 Monitoring und Berichterstattung

- **Performance-Monitoring:** Überwachung der KI-Qualität
- **Bias-Detection:** Regelmäßige Überprüfung auf Verzerrungen

9. Versionierung und Aktualisierung

9.1 Aktuelle Version

- **Dokumentversion:** 1.0
- **Systemversion:** 3.5
- **Letzte Aktualisierung:** 15.07.2025

9.2 Änderungshistorie

Version	Datum	Änderung	Verantwortlich
1.0	15.07.2025	Initiale Dokumentation	Lukas Kauderer (CEO)

10. Kontaktinformationen

10.1 Ansprechpartner

- **Technische Fragen:** Lucas Bäuerle, lucas@licili.de
- **Compliance-Fragen:** Lukas Kauderer, lukas@licili.de
- **Datenschutz:** Niklas Hanitsch, dsb@secjur.com